|  |  |
| --- | --- |
| **PROJECT SPECIFICATION** | |
| **Data Integration Pipelines for NYC Payroll Data Analytics** | |
| **Linked Services** |  |
| CRITERIA | MEETS SPECIFICATIONS |
| The learner will be able to create a Linked Service to configure a connection to Azure Data Lake Gen2 containing employee and payroll data | A Linked Service object is present in the data pipeline repository of type "AzureBlobFS" that configures a connection to Azure Data Lake Gen2 containing payroll data. |
| The learner will be able to create a Linked Service in Azure Data Factory from Azure SQL Database to configure a connection to Azure SQL Database containing employee and payroll data | A Linked Service object is present in the data pipeline repository of type "AzureSQLDatabase" that configures a connection to Azure SQL Database containing employee and payroll data. |
| The learner will be able to create a Linked Service to configure a connection to Azure Synapse Analytics | A Linked Service object is present in the data pipeline repository of type "AzureSqlDW" that configures a connection to Azure Synapse Analytics to land the data for the NYC Synapse Data Warehouse analytics platform |
| **Datasets** |  |
| CRITERIA | MEETS SPECIFICATIONS |
| The learner will be able to create datasets to provide views of employee and payroll data in Azure Data Lake Gen2. | Multiple dataset objects are present in the data pipeline repository of type "AzureBlobFSLocation" with schemas from "AgencyMaster.csv", "TitleMaster.csv", "nycpayroll\_2021.csv", and "EmpMaster.csv" to provide datasets for data views from Azure Data Lake Gen2. |
| The learner will be able to create a dataset to provide a view of employee and payroll data in the Azure SQL DB table. | A dataset object is present in the data pipeline repository of type "AzureSqlTable" with schemas from the NYC Payroll Data SQL DB table to provide a dataset for a data view. |
| The learner will be able to create datasets to provide views of data to the Synapse Data Warehouse platform | Multiple datasets are present in the project repository dataset directory of type "AzureSqlDWTable" to land the data from Azure Data Lake Gen2 into the Synapse Data Warehouse platform. |
| **Data Flows** |  |
| CRITERIA | MEETS SPECIFICATIONS |
| The learner will be able to create data flows to aggregate payroll data from Azure SQL DB and NYC Payroll history files to the Synapse Data Warehouse. | A Dataflow object is present in the data pipeline repository of type "MappingDataFlow" with a union to create a derived aggregated column with the total amount paid to an employee (TotalPaid = RegularGrossPaid + TotalOTPaid + TotalOtherPay). The data sources for this aggregate column should be the data from Azure SQL DB table and the historical NYC payroll data. |
| The learner should be able to create data flows to move data from one data storage system to another. | Multiple Dataflow objects are present in the data pipeline repository of type "MappingDataFlow". Data flows should map data in datasets from Azure Data Lake Gen2 to Azure SQL DB. Data flows should map data from Azure SQL DB and Data Lake Gen2 to move it to the Synapse Analytics platform. |
| **Pipeline** |  |
| CRITERIA | MEETS SPECIFICATIONS |
| The student will be able to create a data pipeline containing Dataflow activities. | Multiple pipeline objects are present in the data pipeline repository with activities of type "ExecuteDataFlow" in the pipeline directory which contain Dataflow objects. |
| The learner should be able to trigger a pipeline and execute the Dataflows in it. | A screenshot is present showing a successful pipeline execution in Azure Data Factory |